

ABSTRACT

A cache replacement system and method for changing the number of cached copies of segments of a media clip in response to a rank change for the media clip. Whenever a rank change for a media clip is detected at an origin server, the rank change information is distributed to a plurality of proxy servers which collectively make up a loosely coupled distributed cache. Upon receiving the rank change information at each proxy server, the caching probabilities for the segments of the media clip are re-computed using the rank change information to determine which segments of the clip to store or discard, thereby forming a new cache layout for the clip at each proxy server. Although each proxy server makes a local determination regarding which segments of the clip to store or discard, segments are neither added or deleted to actually build the new cache layout until such time as client requests for segments of the media clip are received at the proxy servers. That is, the construction of the new cache layout is deferred or postponed until client requests arrive at the proxy servers which initiate the operations of lazy caching and token exchange to construct the new cache layout. Upon receiving client requests, the construction of the new cache layout occurs on a segment-by-segment basis.